Electrical & Electronic Engineering Research

Innovative of Archives Management in China Based on the Application of Big Data

Caiming Zhang[*1], Wei Wang[*2], Xiaoyi Wen[3], Taiping Pan[4], Xinglong Ning[5], Menghan Yang[6], Jiantao Wang[7], Peng Wang[8]

[*134]China University of Labor Relations, Beijing, China

[25678]Gansu Jiyoupin Network Technology Co. LTD

*Corresponding author, e-mail: zhangcaiming@cule.edu.cn

Abstract: This article is fully combed the related literature, the application of big data in the archive management innovation with a thorough system of literature analysis, and obtained from file data acquisition, data transmission, data storage, data archives utilization, staff training and safety archives construction aspects of the need to create a file management system based on big data applications. And Finally, this article analyzes the archival management innovation practice and its effect based on big data application in China, and draws the corresponding value conclusion.

Key words: Big Data Application; Archives Management Innovation; Chinese Archives Management

1 Introduction

With the continuous improvement of information technology, using computer and network technology to digitize document management has become an important part of the reform of many enterprises, government agencies and social organizations. For the processing of transfer the paper documents into the digital documents, the previous implementation was too rough and discrete. With the emergence of OCR technology for Chinese characters, the documents could be scanned for character recognition and then stored in the computer. The exploration and practice in this field has achieved great results. But there was also some obvious shortcomings, even the efficiency of computer processing paper documents has been significantly

improved, the management of the whole input process is not a complete set of information system and the monitoring and coordination of the whole input process is difficult to guarantee. Because there are many kinds of archives and their management systems with different functions design and implementation. Meanwhile the standard of data specification is no unified in the implementation of archives management systems. Paper documents sorting is very time-consuming and laborious which requires a higher level of knowledge and management experience. Sometimes a large decluttering job can take dozens of professionals or dozens of days or more to complete, and it's easy to slip up in the process. In view of this, many enterprises and departments in the treatment of paper documents, often be fall into and "only archive, not collate" situation.

Information technology in the application of archives management is very necessary and important. In the process of and collection of files by the using of the information technology can improve the quality and level of archives management. And also, a good job of software operation, avoiding mistakes and the scientific nature of archival data should be confirmed in the useing of computer technology to manage the archives.

2 Reviews

Liu Yong et al. (2017) analyzed and demonstrated the archival work under the background of big data from the organic combination of platform, management, technology, resources, services and talents, and endowed the archival department with the role positioning of management center, data supervision center and information service center. According to Hou Jia (2013), the construction of smart archives management is faced with information challenges, which provides unprecedented opportunities for archival departments, and proves the irreplaceable important role of archival work from four aspects: "dead archives" to "living information" with the interconnection of archival information, the in-depth development of archival information, and the consciousness of serving the overall situation. In order to ensure the truth and security of the "front end" of basic data, archives departments must integrate the relevant original information resources scattered in various systems within the government organization, and build a unified and authoritative information resource system and service platform to realize information sharing.

In the era of big data, archival data has the characteristics of risk clustering, comprehensive crossover, dynamic ubiquity and hidden relevance. Archival data are not only stored on paper carrier, but also in the archival information system characterized by binary. The program problems and technical vulnerabilities of the archival information system makes the archival data more easily deformed and tampered (Jin Bo and Yang Peng.2020). The training system of archival data security managers is not perfect and durable, and the archival data security talent team with wide knowledge, interdisciplinary and high skills has not yet been built (Qin Qiaoyun, Zhou Feng, Yang Zhiyong,2017). The application of big data in archive management innovation is urgent and has great significance in the following aspects:

2.1 Archive data collection

The archive collection is the first step for archive data to enter the system. Ensuring the security of archive data in the collection stage lays the foundation for archive data security governance. Firstly, the format of archival data should be standardized. Because of the different sources and structures of archival data, there are mane structured, semi-structured and unstructured forms. It is important to integrate and preprocess the data sources in the archival data processing(Shao Qifeng, Jin Chee-qing, Zhang Zhao, Qian Weining, Zhou

Ao-ying,2018) to prepare for the subsequent provision of high-quality archives. And it would ensure the data compliance, consistency and legitimacy (Feng Dengguo, Zhang Min, Li Hao,2014). Secondly, to ensure the authenticity of archival data into the system, one of the risks of archival data for malicious creation of "fake files". Archival data traceability technology, blockchain technology, time stamp and other technologies can be effectively protected the authenticity of archival data from tampering (Yu Yarong, Zhang Zhaoyu,2020)

2.2 Archival data transmission

Because of its high value and large quantity, the archival data will be subject to data security threats from different degrees and aspects in the transmission process of network collection and utilization. Firstly, the archival data is easy to be attacked by hackers in the process of transmission. The archival data encryption technology and archival data anonymity technology are adopted to protect the security of archival data during transmission. The encryption technology of archival data uses symmetric encryption and asymmetric encryption to encrypt and protect key archival data, so as to ensure the security of archival data in transmission. At the same time, in the process of archival data transmission, the relevant network protocol is signed to coordinate the relationship between parties, and the relevant security responsibilities of the responsible party are clear. Secondly, many data files in the system is dependent to the environment of system processing, and the archival data migration will be missing or unable to read file data, solve the problem of file data heterogeneous, from the national strategic height build file data storage platform, in order to solve the file data platform heterogeneous problem (Qin Qiaoyun, Zhou Feng zhi-yong Yang, 2017).

2.3 Archival data storage

While the archival data is stored in the archival information system, the function of the information system affects the archival data security directly. It needs to improve the safe operation and the management system of the archival information system, to improve the prevention and control ability, the authority setting and management level of archival storage system.The archival information system manages the whole process from collection to destruction of archival data. First of all, the software and hardware equipment of archival information system should be strengthen and updated constantly, the security prevention and control ability of software equipment should be improved constantly, and the system vulnerability problem should be reduced constantly. Secondly, the archival information system risk and its crisis response ability should be improved, the anti-intrusion detection technology should be enhanced its response ability constantly with the firewall technology. Firewall technology can effectively block the illegal access and private the data encryption and prevent from the Internet illegal invasion. To improve the fine granularity of access to the archives information system and improve the ability to repair loopholes Continuously to identify malicious attacks and monitoring data process will determine the malicious invasion of computer and network behavior (Liu Xuan, Dong Xin Luna, Ooi Beng Chin,2013).

2.4 Archival data utilization

The archival utilization is the fundamental purpose of archival work which could be realized to the greatest extent. Zhi-yong Yang, Zhou Feng (2016) discusses the wisdom city under the background of archives service four elements in the process of combination and interaction with different emphases of four different kinds of

archives service mode, so as the subject, object and content and archives service to realize the diversification and individuation of archives service, and achieve more accurate efficient and convenient analysis of archive information dissemination. Lyu Yanbing (2016) put forward the view of archival data resources based on the construction of open data, and demonstrated three key areas of future digital archival resources: open data, credit data and crowdsourced data. The archival data utilization is the starting point and foothold of archival data management. The core of archival data in the era of big data is not who owns the data, but who takes these archival data to do what, and what kind of value can be created. The most fundamental significance of archival data management is to realize the deep layer efficiency of archival data, protect the data and ensure the utilization. So, to establish the security sharing mechanism in the process of archival data utilization, to establish the archival data alliance sharing platform is a good alternative. For example, the Australian National Data Service Center has combined the data information resources of more than 100 Australian research institutions, governments and universities, with the integrating the data resources and improving the process of data development and sharing successfully (Chen Wenjie, Cai Lizhi,2016). The access control mechanisms of archival information system with a corresponding strength or granularity according to compliance requirements should be established, including the identity authorization and data traces, and reasonably to limit the access scope of visitors. In the process of using archival data, identity identification, access control rights and other issues should be carried out to determine which archival data can be used and which people can access which data (Liu Yuenan,2020). Based on the different subjects, the different archival data, and the different network systems, the security protection system need to combine to the archival data permission control with archival user permission control to protect the key information of archival data.

2.5 Archives professionals training

Archivists are the keepers and guardians of archival information in various units and periods. The realization of archival value needs the transmission of archivists. Therefore, with the development and innovation of archival work under the background of big data, the archivists are the key factors that determine the development degree of archival work. Along with the Internet of things, big data and cloud computing, new media, and the rapid development of information technology and application, the knowledge and wisdom, the good information literacy is not only become the necessary quality of archivists, but also become the development of archive work innovation technical support. Gao Qixiang (2017) starting from the necessity of information literacy promotion, introduced from national policies and industry norms for the archival institutions evaluation mechanism, the personal professional identity: personal accomplishment (that is, the service quality and media literacy) and the information ability (i.e., data management and data mining skills). Zhiyong Yang and Meirong Fei (2017) believe that archivists should be "professional" and "knowledgeable" in order to meet the needs of archival work in the digital age and smart cities. Wang Shang (2017) believes that the construction of archival talents should include the selection, evaluation, use, education and retention of personnel, so as to grasp the latest changes in the field of archives and continuously expand the scope of knowledge. Tang Yizhi (2014) starting from the repositioning of archives management, thinks the archivists should learn, use and master the new technology especial the new information technology. Xue Jinling (2012) discusses the construction of archival talents from the aspects of professional quality, information awareness, professional knowledge, information technology and foreign language proficiency.

2.6 Archive security construction

The archival security refers to the taking effective protection measures to protect archival entities and contents from natural disasters or man-made damage. Jin Zhizhi (2016) elaborated on the category of archival information security (including transmission security and media security), including "strengthening relevant security business training", "building a security monitoring system of Internet + Internet of Things", "Cloud grading, dividing sharing and security index", "timely security backup", The four aspects of risk reduction index demonstrate the promotion strategy of archival information security. Zhai Fei (2016) analyzed the practice in the security risks of archives information system, found the five aspects to explore the archives security including the establishing and improving of the safety management system, the scientific predictions of security risks, the ensuring of long-term effective information storage, the improving of information security awareness, the detailed specification of information security standard system, in order to take corresponding prevention measures in practice, and puts forward the security strategies of archives information under the network environment from seven aspects such as the implementation of network security evaluation.

3 Archives management system

It includes two sub-systems, which is the front-end file collection system and the archive collection system.

3.1 Front-end file collection system

The front-end file acquisition system is a newly developed digital file processing program aiming at the requirements of file attributes, which can comprehensively cover all kinds of files. Combined with the self-developed OCR region automatic recognition technology, it can quickly and accurately identify all kinds of description items needed by catalog data, and generate electronic files intelligently. Instead of the traditional "manual catalogue description - scanning - connecting" digital processing flow.

3.2 Archive collection system

The archives collection system, broken the traditional filing methods, restructuring the files collect 15 steps, build the "document scanning, sorting, online online review" file finished the three-step model, which would be combined with improved TextRank method and digital watermark technology, intelligent generated page number, package number, file number, chapter file, archive file directory, the reference appendix table, etc., Completely replacement of manual input , to achieve the collection of archives automation and networking.

4 Application of Archives management system

The archives digitization process to avoid the rework and waste of resources, and not only save the storage cost and space, but also be very convenient and fast, to avoid the repeated printing materials and cause the waste of paper and personnel. In the process of archives management, the efficiency of archives collection and collation will be significantly improved with the application of Archives management system. In the past,

a lot of manual working on the sorting, collection, classification, records, search and management, not only cost a large amount of human resources, also reduced the accuracy of the records management work, and people's energy will be greatly depleted. However, after the application of archival information management system, the limitation of this management work is easily broken. More perfect archival management methods and system have been applied, and relevant personnel can conduct detailed processing, differentiation, screening and screening of archival information, and the archives management work has be sustainability. The archives management system establish a classification management module, and could adjust the instruction at any time according to the situation.

5 Conclusions

The innovative practice of archives management system in China based on the application of big data can produce huge value and benefit.

5.1 Saving the costs and reduce rework

In the past, the extensive model has been used to increase office staff and office costs as the only means to solve this problem, and it caused a waste of resources, large amount of rework, resulting in a significant increase in management costs. The archives digitization process could avoid the rework and waste of resources. And the digital archives management system make traditional paper as the carrier of the archives information object into a machine readable file, not only save the storage cost and space, but also is very convenient and fast, to avoid the repeated printing materials and cause the waste of paper and personnel.

5.2 Making the archival work be sustainable

In the process of archives management, the efficiency of archives collection and collation will be significantly improved. With the application of archival information management technology system, the limitation of manual management work is easily broken. More perfect archival management methods and technology systems have been applied, and relevant personnel can conduct detailed processing, differentiation, screening and screening of archival information. In the application of information technology, people can find a more rapid way to work, so as to reduce the probability of data loss, data information integrated management, so that the archives management work has been sustainability.

5.3 Establishing a standardized archives management system

The classification management mode of archives based on tagging application has been changed. The traditional file management includes documents and archives, infrastructure data, personnel files, unit development history, etc., But with the application of archival information management technology system , a classification management module and system will be easy to be established. With the use of computer technology to set the instructions, the electronic archives can be divided into many categories according to the unit production, finance, personnel, administrative management, and establish a perfect and standardized file management system. The efficiency of archive collection and collation will be significantly improved by the application of archives management system.

References

Jin bo, Yangpeng. (2020). Analysis of archival data security governance strategy in the era of big data. *Information science*, 38(9):30-35.

Wang Shiwei. (2017).Smart society is a new realm of smart library development. Library journal, 36(12):9-13.

Qin Qiaoyun, Zhou Feng, Yang Zhiyong. (2017). Research on information security of digital Archives in big data environment. *Beijing Archives,* (6):18-21.

Shao Qifeng, Jin Cheqing, Zhang Zhao, Qian Weining, (2018). Zhou Aoying. Blockchain technology: architecture and progress. *Journal of computing,* 41(5):969-988.

Feng Dengguo, Zhang Min, Li Hao. (2014). Big data security and privacy protection. *Chinese journal of computers*, 37(1):246-258.

Yu Yarong, Zhang Zhaoyu. (2020). Design of electronic archive evidence collection and verification scheme based on trusted timestamp service. *Archives Management*, (1):66-68.

Liu Xuan, Dong Xin Luna, Ooi Beng Chin, (2013). Srivastava Divesh.Online data fusion//Proceedings of the 39th International Conference on Very Large Data Bases (VLDB '2013). Trento, Italy, 97-108.

Chen Wenjie, Cai Lizhi. (2016). Big data security and its evaluation. *Computer applications and software*, 33(4):34-38+71.

Wu Jiaqi, Chen Xiaoling. (2015). Research on Co-construction and sharing Mode of Regional Archival Information Resources under the background of smart City. *Archives Management*, 1:33 ~ 35

Lv Yan ice. (2016). Digital Archive Resources in Smart City Framework. Journal of zhejiang archives, 1:24 ~ 26

Yang Zhiyong, Zhou Feng. (2016). Research on Archival Information Service Mode for Smart City. Bulletin of archives science, 4:44 ~ 49

Jin Zhizhi. (2016). Information Security Promotion Strategy of Smart Archives. Management Observation, 6:46 ~ 48

Zhai Fei. (2015). Research on Digital Archive Information Security Guarantee System. Office Business, 2016, 22:59 ~ 60

Wang Yihan. (2015). Research on the Security System of Archive Information Management in Network Environment. Technology development of enterprise, 34 (17) : 123 ~ 124